DOCUMENT RESUME

ED 202 871 TM 810 290

AUTHOR King, Lynda A.
TITLE Generalizability Approach to Validity of Gender-Related Constructs.
PUB DATE Apr 81
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Analysis of Variance; *Attitudes; Mathematical Models; Peer Evaluation; Rating Scales; Self Evaluation (Individuals); *Sex Role; *Test Validity
IDENTIFIERS Constructs; *Generalizability Theory; *Multitrait Multimethod Techniques

ABSTRACT

The assessment of the multitrait-multimethod matrix is placed in the context of generalizability theory for the purpose of demonstrating the convergent and discriminant validity of gender-related constructs. Once convergent validity is established with the traditional generalizability/reliability approach, a two-facet (traits and methods) generalizability design is proposed, with the levels of the trait facet comprised of scores on three separate measures. In this latter design, emphasis is placed on examination of the subject x trait and the subject x method interactions, with equations presented to aid in interpreting the degree to which constructs account for unique variance in the measurement of sex-role attitudes. (Author/RL)

Generalizability Approach to Validity of Gender-Related Constructs

Lynda A. King

Central Michigan University

Portion of a symposium entitled Validity Issues in Gender-Related Research: Measurement and Design, presented at the annual meeting of the American Educational Research Association, Los Angeles, California, April 1981.

The author would like to express her gratitude to the Faculty Research and Creative Endeavors Committee, Central Michigan University, for its support for this presentation.

Generalizability Approach to Validity of Gender-Related Constructs

Campbell and Fiske (1959) introduced the multitrait-multimethod approach to construct validation, with its emphasis upon the examination of convergent and discriminant validity as evidenced in the multitrait-multimethod matrix. While Campbell and Fiske did not provide any specific analytic technique for evaluating convergent and discriminant validity, they listed four general criteria by which to assess the matrix: (a) The correlations between different methods on the same trait should be significantly different from zero. (b) The correlations between different methods on the same trait should be higher than the correlations between different methods on different traits. (c) The correlations between different methods on the same trait should be higher than the correlations between different traits using the same method. (d) The pattern of correlations among traits should be the same within and between methods.

Subsequent to the publication of Campbell and Fiske's (1959) classic work, several researchers suggested methods for analytically determining the validity of a construct using the multitrait-multimethod paradigm. Tucker (1963) and Jackson (1969) have used factor analytic procedures to assess the multitrait-multimethod matrix. Stanley (1961) proposed the use of an analysis of variance methodology; this approach has been more recently used by, among other, Boruch, Larkin, Wolins, and MacKinney (1970) and Kavanagh, MacKinney, and Wolins (1971).

While the present paper will use an analysis of variance approach to construct validation, it will attempt to place the ass·ssment of the multitrait-multimethod matrix in the context of generalizability theory

(Cronbach, Gleser, Nanda, & Rajaratnam, 1972). Generalizability theory is usually considered a multifaceted reliability approach; however, as recommended by Mitchell (1979), it has other potential applications. In the present paper it will provide the theoretical framework to examine the validity of constructs used in gender-related research. A further point should be made: The generalizability perspective as explored in this paper is not just that of the "test developer" interested in the precision of measurement, but rather that of the theoretician or "construct developer" interested in both the presence and absence of convergence. Because of its emphasis on the use of more complex designs, the generalizability perspective enables the researcher to gain a more accurate picture of the construct in question -- what it is and what it is not.

Generalizability theory views a behavior as occurring at the junction of an infinite number of conditions or circumstances. Conditions might include the occasion on which a measure was taken, the method by which the measure was obtained, or the source of information from which data were obtained. Cronbach et al. (1972) defined the total set of the conditions that comprise a measurement situation as the universe of admissible observations, and sets or classes of conditions (e.g., occasions, raters, items, etc.) as facets. Generally, it is expected that observations of subjects across levels of a facet covary such that the interactions of facets with subjects (the objects of observation) are seen to comprise measurement error.

Thus, returning to the multitrait-multimethod analysis, the common generalizability study for a single trait could provide a monotrait-monomethod element and a monotrait-heteromethod element for the multitrait-multimethod matrix. Both of these elements would be representative of convergent validity or the extent to which the same trait is being consistently measured. However, if one can postulate another facet representing the construct of interest and other constructs which are expected not to covary with it, then one is able to stretch the generalizability model to fit the full multitrait-multimethod approach by introducing a source of evidence for discriminant validity. Therefore, what was previously called a universe of admissible observations has been somewhat changed in that the multitrait model calls for consideration of inadmissible observations. In other words, covariance among traits is not desired.

Application to Gender-Related Constructs

In a review of measures of gender-related constructs, Beere (1979) noted confusion in the definitions of constructs being measured. According to Beere, some constructs were merely insufficiently defined and thus weak item sampling resulted. In other cases, variables designated with the same name appeared to be measured by very different items, while variables designated with different names appeared to be measured by very similar items. The overall result seems to be a profusion of inadequately conceptualized measures with little data concerning the validity of the various constructs that they represent.

Guided by Beere's (1979) explication of the deficiencies of previous measures, King, Beere, King, and Beere (1981) proposed the construct of "sex-role egalitarianism" and developed a measure of the construct using a rational scale development approach (Jackson, 1971). Sex-role egalitarianism was defined as an attitude which causes one to respond to another individual independently of that other individual's sex. One who possesses this attitude believes that the sex of a person should not influence the perception of that person's abilities or the determination of that person's rights, obligations, and opportunities. Consequently, a sex-role egalitarian does not discriminate against or relate differentially to another on the basis of the other's sex. Five domains of "adult living" in which the construct was hypothesized to manifest itself were defined and items designed to tap these domains were selected, with an emphasis on internal consistency.

A rather unique aspect of the construct as defined by King et al. (1981) was its purposeful disregard of the sex of the individual serving as the attitude object. That is, it was noted by the construct developers that previous instruments measuring sex-role attitudes had frequently focused on attitudes toward females in nontraditional settings (e.g., the popular Attitudes Toward Women Scale, Spence & Helmreich 1972). Furthermore, even when the developers of existing scales suggested that their instruments were designed to measure "sexual equality" with no reference to which sex's equality, an examination of the actual items included in the instruments revealed that they were almost exclusively concerned with females in nontraditional sex roles and not with males in nontraditional sex roles (e.g., the Sex Role Survey, MacDonald, 1974).

To clarify the validity of the sex-role egalitarianism construct, a sequence of designs could be developed in which the discriminant and convergent validity of the construct can be evaluated. In particular, it seems important to determine the extent to which the notion of egalitarianism as defined here is different from merely "attitudes toward women" or "attitudes toward the ability of women to equal men in traditionally male roles." The construct is defined as going beyond such definitions and should therefore manifest limited convergence with other instruments measuring sex-role attitudes from these more restricted perspectives.

Designs for Exploring Construct Validity

For clarity and ease of presentation, let us assume that the validity study involves three specific instruments: (a) the Sex-Role Egalitarianism Scale (SRE), purported to measure the extent to which one responds to another independent of the other's sex; (b) the Attitudes Toward Women Scale (AWS), purported to measure attitudes toward the rights and roles of women in contemporary society; and (c) the Sex Role Survey (SRS), purported to measure attitudes regarding equality between the sexes. Assume further that the major goal is to establish the validity of the SRE construct, which is believed to account for variability unaccounted for by the other variables.

Perhaps the preliminary task for the construct developers is to demonstrate that the measure of interest is consistent across various conditions. In this initial phase, the competing constructs are ignored in favor of attention to whether or not measurement with the instrument of interest is consistent across items, forms (if multiple forms are available), occasions, etc. The design and analysis in this case focus on the definition of the universe of admissible observations, and the aim is to maximize covariance of subjects across items, forms, etc. In generalizability terms, for example, one might employ a fully-crossed two-facet (subjects x items x occasions) design and examine the relative size of the resulting variance components. The expectation is that the subject component of variance be large relative to components representing interactions of facets with subjects. This type of design emphasizes reliability or accuracy of measurement, answering the question: "Can the construct be dependably measured?" From a multitrait-multimethod perspective, the construct developers are gaining information regarding the monotrait-heteromethod and the monotrait-monomethod elements.

The establishment of dependability in such a manner would next call for consideration of the heterotrait-heteromethod and heterotrait-monomethod elements. Inclusion of a trait facet, as previously noted, introduces the notion of a universe of inadmissible observations. Scores on the SRE scale would now be joined by scores on the AWS and SRS to form levels of the trait facet. A potential design in this instance would be a fully-crossed two-facet design, traits and methods crossed with subjects and with one another. For example, each of the three traits or constructs could conceivably be measured by two methods. Methods could represent the source of information, for example, peer and self-ratings. If alternate forms existed for all three instruments, then alternate forms could be designated as levels of the method facet.

Table 1 and Figure 1 present components of variance for a fully-crossed, random effects, two-facet (subjects x traits x methods) design, with traits being SRE, AWS, and SRS, and methods being self-ratings and some form of peer ratings. Contrary to the previous emphasis on accuracy of a given construct, the focus now is on demonstrating convergence across methods for each single trait and discrimination among traits regardless of method. The desire for a large subject component of variance is now replaced by a concern for the size and nature of interaction components of variance.

Given the present design, the interaction components of variance that are most salient to determining the validity of the SRE construct are that representing the subject x trait interaction and that representing the subject x method interaction. Each reflects a residual covariation used as evidence for construct validity. Equations presented by Stanley (1961) provide guidance to the construct developers in assessing convergent and discriminant validity.

Let us first consider the interaction between subjects and traits. Under the random effects model, this interaction is the weighted sum of two variance components, σ^2_{stm} and σ^2_{st}. Stanley (1961) demonstrated that this mean square was directly related to the magnitude of the mean of the covariances within traits across methods (A), inversely related to the mean of the covariances within methods across traits (B), and inversely related to the mean of the covariances across both methods and traits (C). Of particular interest are A and C. In order for there to be discriminant validity, that is, that the traits not covary, it is important that the methods do covary within each trait. In other words, if measures on the same trait using different methods disagree across subjects or if subjects are differentially rated on the same trait as a function of the method, discrimination among traits cannot occur. It is advisable that A be larger than D; in Campbell and Fiske's (1959) terminology, heterotrait-heteromethod coefficients should have lower values than monotrait-heteromethod coefficients. The difference between these two covariance estimates is equal to the variance component σ^2_{st} and can be computed directly by the formula

$$\sigma^2_{st} = \frac{1}{N_m} (MS_{st} - MS_{stm})$$

where N_m is the number of levels of the method facet.

In like manner, the subject x method mean square, under the random effects model, can be shown to be a weighted sum of two variance components, σ^2_{stm} and σ^2_{sm}, and related to the magnitudes of covariances A, B, and C above. It is desirable that the covariance within methods across traits (B) be small since a large value for B would suggest method bias. In contrast, a low value for B would demonstrate discriminability among traits even at the level of a single method, suggestive of heterotrait-monomethod coefficients. Under optimal conditions for construct validity, the pattern of heterotrait-heteromethod and heterotrait-monomethod correlations should be similar. Hence, B and C should have relatively similar values. B - C can be computed as $\sigma^2_{sm} = \frac{1}{N_t}(MS_{sm} - MS_{stm})$, where N_t is the number of levels of the trait facet.

Finally, it should be noted that the model presented here is sufficiently flexible to incorporate additional dimensions or facets into the multitrait-multimethod paradigm. For example, if multiple administrations of the several instruments are feasible, then one is able to determine covariance not only over methods, but over the separate occasions when these methods are used. In this manner, information about the stability of the construct could be derived; although traditionally considered as "reliability" data, it has added meaning when placed in the context of multiple traits.

The above strategy for exploring the validity of a construct is certainly not restricted to gender-related constructs. However, because the sex-role literature is growing at a rapid rate, and because this growth is necessarily accompanied by the need for meaningful, useful constructs measured by reliable and valid instruments, it is appropriate and timely that those concerned with this field of study begin to systematically "sort out" the variables of constructs that are its foundation. When one speaks of "attitudes toward women," is one really speaking of the cognitive, emotional, and behavioral reaction of persons toward women in nontraditional roles? When one speaks of "sexual equality," is one thinking, consciously or unconsciously, unindirectionally in that the construct implies "the ability of women to equal men," or is equality bidirectional in that it does (or should?) incorporate the notion of men's ability and/or freedom to equal women in nontraditional roles? The SRE construct (King, et al., 1981) offers the belief that bidirectional equality is a useful basis for studying sex-role behavior. The multitrait-multimethou/generalizability approach discussed in this paper is merely one mechanism to assess its unique contribution.

References

Beere, C. A. Women and women's issues: A handbook of tests and measures. San Francisco: Jossey-Bass, 1979.

Boruch, R. F., Larkin, J. D., Wolins, L., & MacKinney, A, C. Alternative methods of analysis: Multitrait-multimethod data. Educational and Psychological Measurement, 1970, 30, 833-853.

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: John Wiley, 1972.

Jackson, D. N. The dynamics of structured personality tests: 1971. Psychological Bulletin, 1971, 78, 229-248.

Jackson, D. N. Multimethod factor analysis in the evaluation of convergent and discriminant validity. Psychological Bulletin, 1969, 72, 30-49.

Kavanaugh, M. J., MacKinney, A. C., & Wolins, L. Issues in managerial performance: Multitrait-multimethod analyses of ratings. Psychological Bulletin, 1971, 75, 34-49.

King, L. A., Beere, D. B., King, D. W., & Beere, C. A. A new measure of sex-role attitudes. Paper presented at the annual meeting of the Midwestern Psychological Association, Detroit, April 1981.

MacDonald, A. P., Jr. Identification and measurement of multidimensional attitudes toward equality between the sexes. Journal of Homosexuality, 1974, 1, 165-182.

Spence, J. T., & Helmreich, R. L. The Attitudes Toward Women Scale: An objective instrument to measure attitudes toward the rights and roles of women in contemporary society. JSAS Catalog of Selected Documents in Psychology, 1972, 2, 66. (Ms. No. 153).

Stanley, J. C. Analysis of unreplicated three-way classifications with applications to rater bias and trait independence. Psychometrika, 1961, 26, 205-219.

Tucker, L. R. Implications of factor analysis of 3-way matrices for measurement of change. In C. W. Harris (Ed.), Problems in measuring change. Madison: U. of Wisconsin Press, 1963.

Table 1

Sources of Variance and Expected Mean Squares

for Fully-Crossed, Random Effects, Two-Facet Design

Source of Variance	Expected Mean Square[a]
s	$\sigma^2_{stm} + N_t\sigma^2_{sm} + N_m\sigma^2_{st} + N_tN_m\sigma^2_s$
t	$\sigma^2_{stm} + N_s\sigma^2_{tm} + N_m\sigma^2_{st} + N_sN_m\sigma^2_t$
m	$\sigma^2_{stm} + N_s\sigma^2_{tm} + N_t\sigma^2_{sm} + N_sN_t\sigma^2_m$
st	$\sigma^2_{stm} + N_m\sigma^2_{st}$
sm	$\sigma^2_{stm} + N_t\sigma^2_{sm}$
tm	$\sigma^2_{stm} + N_s\sigma^2_{tm}$
stm	σ^2_{stm}

a N_s = number of subjects; N_t = number of levels of the trait facet;
N_m = number of levels of the method facet

Figure 1. Schematic Representation of Fully-Crossed, Random Effects, Two-Facet Design